Data Analytics for Materials Science

A.D. (Tony) Rollett, R.A. LeSar (Iowa State Univ.)

Dept. Materials Sci. Eng., Carnegie Mellon University

Linear Algebra (for Regression Analysis)

Lecture 3B

Revised: 8th Feb. 2021

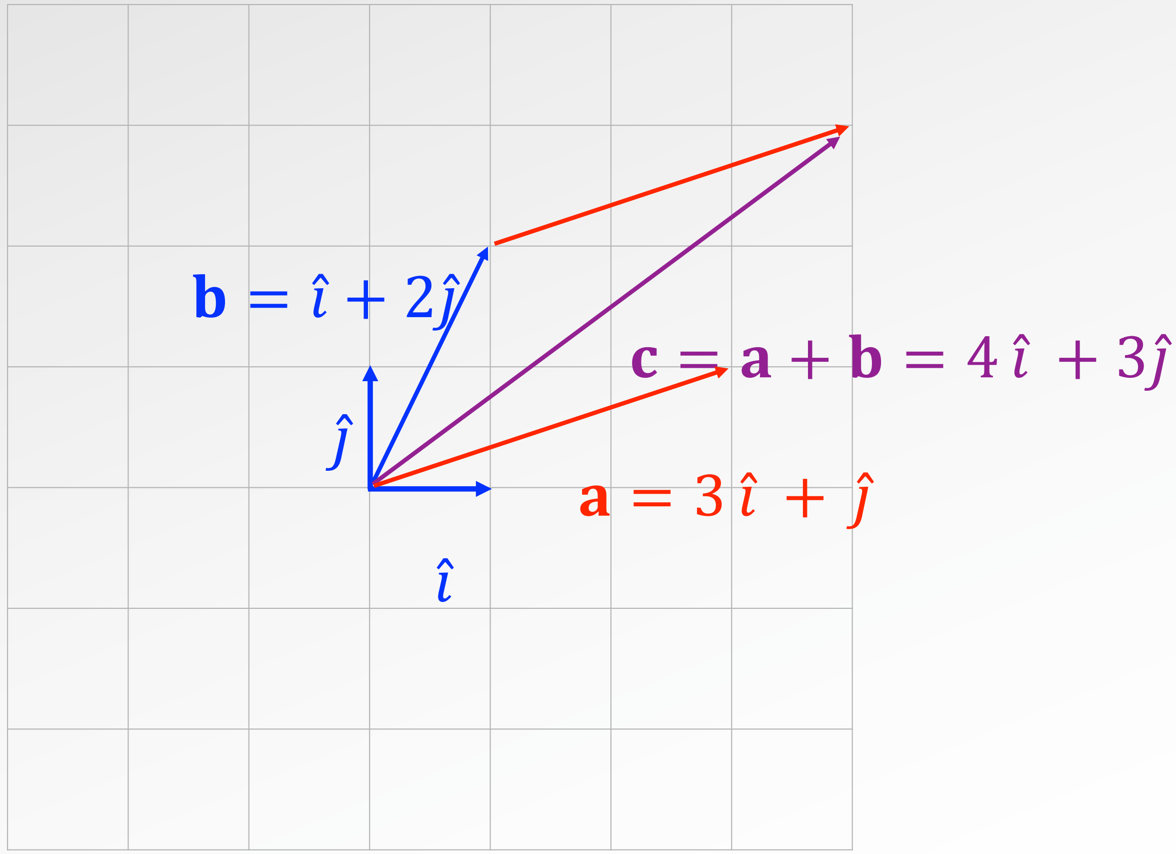
- It is how we use and manipulate (e.g., transform) vectors and matrices to solve problems
- Linear algebra provides the basis for many of the techniques that we will use in this course, most immediately linear regression, but also correlations, principal component analysis and similar techniques, machine learning tools, …

Some background information:

- for a very useful, visual, view of linear algebra, we recommend "Essence of Linear Algebra", which can be accessed at: https://www.3blue1brown.com/essence-of-linear-algebra-page/
- if you want a detailed treatise, try out the MIT courseware site at https://www.youtube.com/playlist?list=PL49CF3715CB9EF31D

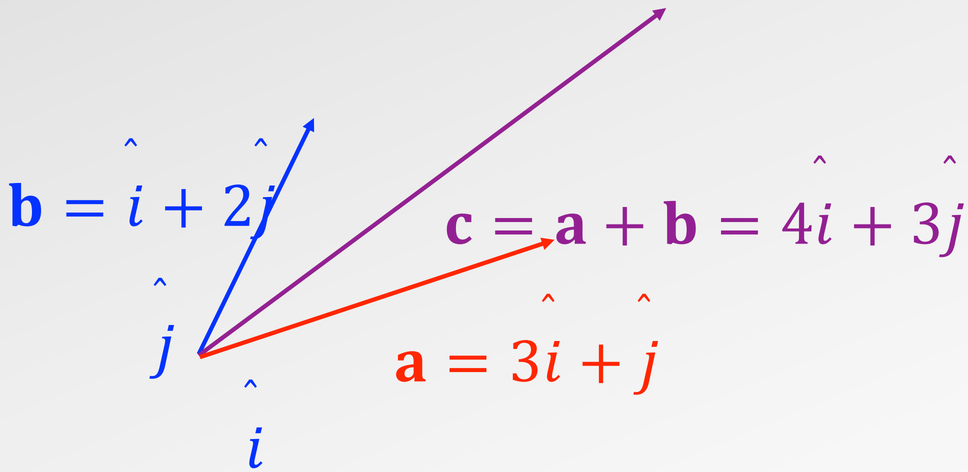
Why linear algebra?

We can write any point in space in terms of $\hat{\imath}$ and $\hat{\jmath}$ — they are the **basis set** *for the space*



Notation: a "hat" indicates a vector of length = 1, called a unit vector

In linear algebra (and computation), we represent a vector as a list of numbers:



$$\mathbf{a} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

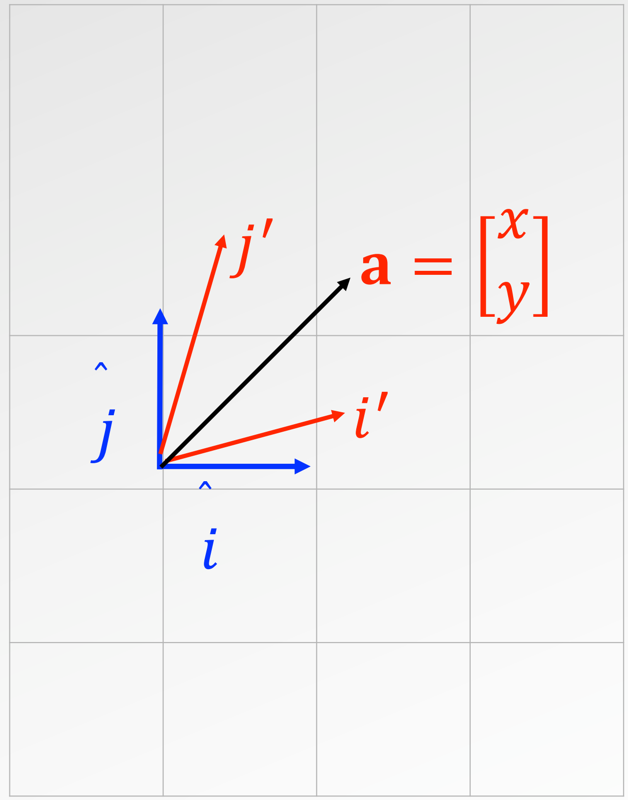
$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

Suppose we want to *transform* the lattice to another set of regular, equally-spaced, points.

We can transform the basis vectors as shown:

$$i' = a\hat{i} + c\hat{j} \qquad j' = b\hat{i} + d\hat{j}$$

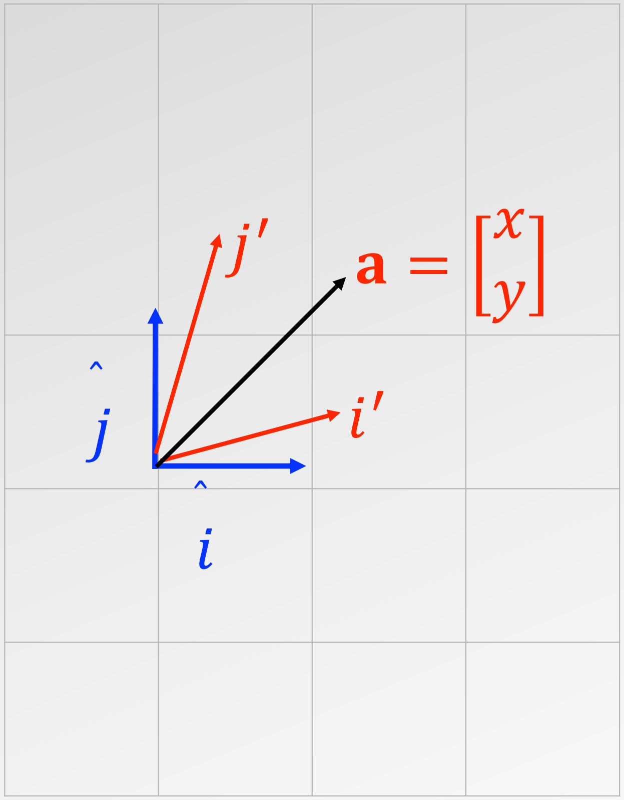
Note: i' and j' are not necessarily of unit length



Question: if $\hat{i}, \hat{j} \rightarrow i', j'$, what happens to a vector $\mathbf{a} = \begin{bmatrix} x \\ y \end{bmatrix}$?

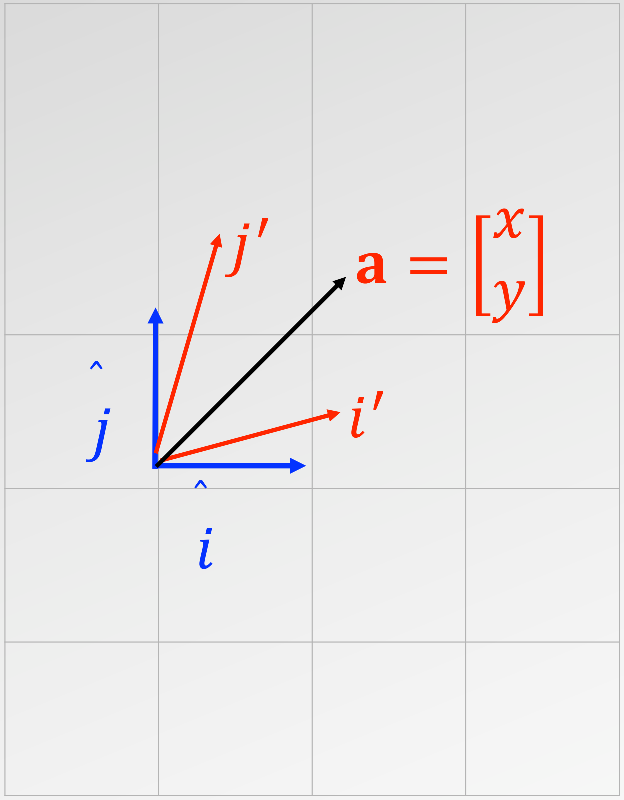
Linear transformations

See Chapter 3 of "Essence of Linear Algebra"



Question: if $\hat{i}, \hat{j} \rightarrow i', j'$, what happens to a vector $\mathbf{a} = \begin{bmatrix} x \\ y \end{bmatrix} = x\hat{i} + y\hat{j}$?

- apply the transformation, $\hat{i} \rightarrow i' = a\hat{i} + c\hat{j}$
- the new position for x under the transformation (x' (is $\mathbf{x}' = x\left(a\hat{i} + c\hat{j}\right)$ $= x\begin{bmatrix} a \\ c \end{bmatrix}$.
- Similarly, $\mathbf{y}' = y\begin{bmatrix} b \\ d \end{bmatrix}$.
- So $\mathbf{a} \rightarrow \mathbf{v}$, with the new vector ($\mathbf{v}$) being $\mathbf{v} = \mathbf{x}' + \mathbf{y}'$



The new vector ($\mathbf{v}$) is

$$\mathbf{v} = \mathbf{x}' + \mathbf{y}' = x\begin{bmatrix} a \\ c \end{bmatrix} + y\begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

We can write this as a *matrix* multiplying a *vector*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}, \text{ or}$$

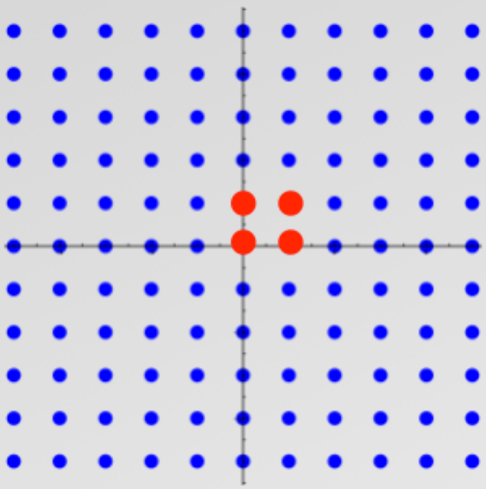
$$\mathbf{A}\mathbf{a} = \mathbf{v} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}, \text{ where}$$

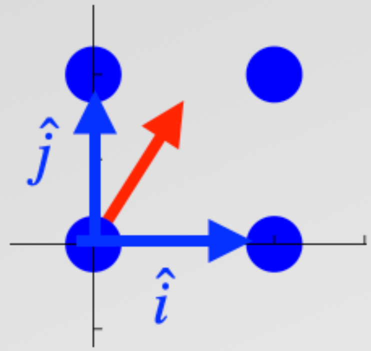
column

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$ row

Linear transformations

See Chapter 3 of "Essence of Linear Algebra"





$$\hat{\imath} \to \hat{\imath}$$

$$\hat{\jmath} \to \frac{1}{2}\hat{\imath} + \frac{\sqrt{3}}{2}\hat{\jmath}$$

$$\hat{\jmath}' = \frac{1}{2}\hat{\imath} + \frac{\sqrt{3}}{2}\hat{\jmath}$$

$$\hat{\jmath}' \cdot \hat{\jmath}' = 1$$

$$\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{bmatrix}$$

Consider a point in the original basis

$$\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$$

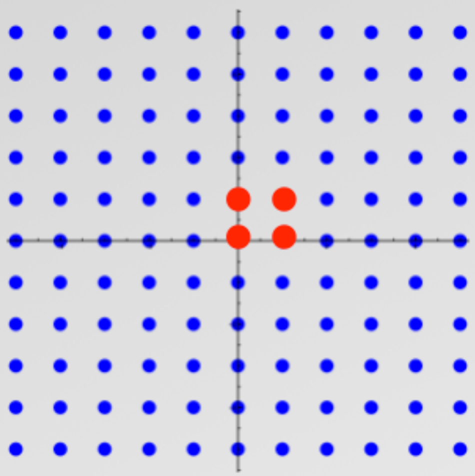
Apply the transformation matrix

$$\mathbf{v}' = \mathbf{A}\mathbf{v} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + \frac{1}{2}y \\ \frac{\sqrt{3}}{2}y \end{bmatrix}$$

For example, if $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, then $\mathbf{v}' = \begin{bmatrix} \frac{3}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$

Linear transformation example

See Chapter 3 of "Essence of Linear Algebra"



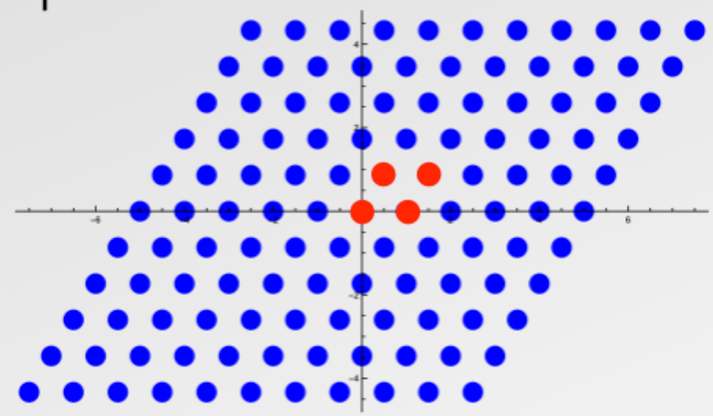
$$\hat{j}' = \frac{1}{2}\hat{i} + \frac{\sqrt{3}}{2}\hat{j}$$

$$\hat{j}' \cdot \hat{j}' = 1$$

$$\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{bmatrix}$$

$$\mathbf{v}' = \mathbf{A}\mathbf{v}$$

Apply transformation to every point in the square lattice:



This matrix *transforms* a square lattice into a triangular lattice in 2D.

Think of matrices as *transformations!*

Linear transformation example

See Chapter 3 of "Essence of Linear Algebra"

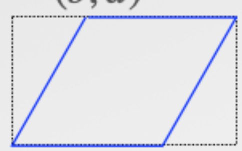
Consider a matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

The determinant of **A** is the area (volume) factor associated with the transformation described by **A**.

It is defined in 2D as $\det A = ad - bc$

$\hat{i} \to a\hat{i}$

$\hat{j} \to b\hat{i} + d\hat{j}$



$\mathbf{A} = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ $\quad \text{Area}(A) = (a+b)d - 2\left(\frac{1}{2}bd\right) = ad = \det A$

For a triangular lattice:

$$\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{bmatrix} \qquad \det A = \frac{\sqrt{3}}{2}$$

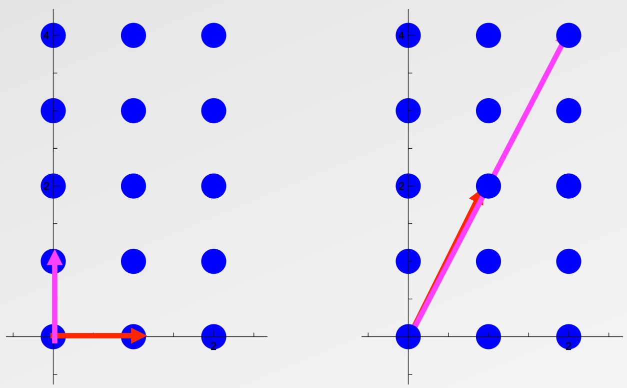
The area of the transformed lattice is about 0.866 times that of the original lattice. Why do we care?

The area is smaller so the density of points is higher.

The determinant of a 2D matrix

See Chapter 6 of "Essence of Linear Algebra"

Consider $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \Rightarrow detA = (1)(4) - (2)(2) = 0.$



The transformation maps the space on to a <u>line</u>, which has zero area, so the determinant is 0.

Note that the vectors for i' and j' are linearly dependent (j' is a simple multiple of i').

How a determinant can be zero

See Chapter 6 of "Essence of Linear Algebra"

Consider the equation $\mathbf{Av} = 0$. $\mathbf{v}$ could be 0 (uninteresting).

If $\mathbf{v}$ is not 0, then think about what the equation says. It says that the transformation maps $\mathbf{v}$ on to a scalar (0), which has zero area.

The only way that that could be true is if the determinant of $\mathbf{A}$ is zero, i.e., $detA = \mathbf{0}$.

We will use this relation in solving for eigenvectors and eigenvalues later on.

If a determinant is zero

See Chapter 6 of "Essence of Linear Algebra"

The determinant of a 3D matrix

Consider $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. What is $detA$?

- In 3D, the determinant is the volume associated with the transformation.
- In nD (n>3), the determinant gives you the "hyper-volume"of the transformation (we will have n>3 in many applications)
- Use a code such as Matlab, Mathematica, … to calculate the determinant for n>2.
- Note: $detA = 0$ if the transformation maps onto a plane or a line

See Chapter 6 of "Essence of Linear Algebra"

Suppose $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$. What is $\mathbf{AB} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix}$?

A times column 1 of **B** is column 1 of the product:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e \\ g \end{bmatrix} = \begin{bmatrix} ae + bg \\ ce + dg \end{bmatrix}.$$

Similarly **A** times column 2 of **B** is column 2 of the product, $\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} f \\ h \end{bmatrix}$ $= \begin{bmatrix} af + bh \\ cf + dh \end{bmatrix}$, so we have

$$\mathbf{AB} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix} = \mathbf{C}$$

Multiplication of two matrices

Consider $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. We define the *inverse* of a **square** matrix $\mathbf{A}^{-1}$ as a matrix that has the property $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

$\mathbf{I}$ is called the *identity matrix* and has the property $\mathbf{Iv} = \mathbf{v}$ or $\mathbf{IA} = \mathbf{A}$.

In 2D, the inverse is given by: $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \mathbf{A}^{-1} = \frac{1}{detA}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

$detA = ad - bc$

Note: if $detA = 0$ (whatever the dimension of the matrix), then no inverse of $\mathbf{A}$ exists.

The inverse of a matrix

See Chapter 7 of "Essence of Linear Algebra"

Consider 3 equations and 3 unknowns — we want to solve for x, y, and z.

$$0x + 1y + 2z = 4$$
$$1x + 2y + 1z = 4$$
$$1x + 0y + 1z = 8$$

We can rewrite this in matrix form as: $\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix}$

The matrix maps the variables (x,y,z) to (4,4,8).

Draw a graph for a problem in 2D and see what it looks like.

See Chapter 7 of "Essence of Linear Algebra"

Using matrices to solve equations

The equation is in the form: $\mathbf{Aa} = \mathbf{v}$, with

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \ \mathbf{a} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \text{ and } \mathbf{v} = \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix}.$$

Multiply $\mathbf{Aa} = \mathbf{v}$ by $\mathbf{A}^{-1}$ from the left (order matters!!), the **a** is found by

$$\mathbf{A}^{-1}\mathbf{Aa} = \mathbf{a} = \mathbf{A}^{-1}\mathbf{v}.$$

As long as the inverse exists (i.e., $\det A \neq 0$), the solution is just $\mathbf{a} = \mathbf{A}^{-1}\mathbf{v}$

(many codes will find the inverse for you).

FYI: $\mathbf{A}^{-1} = \begin{bmatrix} -1 & \frac{1}{2} & \frac{3}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$ and $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 10 \\ -2 \\ -2 \end{bmatrix}.$

See Chapter 7 of "Essence of Linear Algebra"

How do we solve these equations?

Write: $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ and $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

Note the way we have written the elements of the vector and matrix. First index indicates the row, the second the column.

The product $\mathbf{Ab} = \mathbf{c}$ is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_{11}b_1 + a_{12}b_2 + a_{13}b_3 \\ a_{21}b_1 + a_{22}b_2 + a_{23}b_3 \\ a_{31}b_1 + a_{32}b_2 + a_{33}b_3 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

Be careful of "row-major" versus "column-major" numbering of arrays in programming languages

We have $c_i = \sum_{j=1}^{3} a_{ij}b_j = a_{i1}b_1 + a_{i2}b_2 + a_{i3}b_3 \quad i = 1,2,3$

Einstein notation: $c_i = a_{ij}b_j$ (repeated indices are summed)

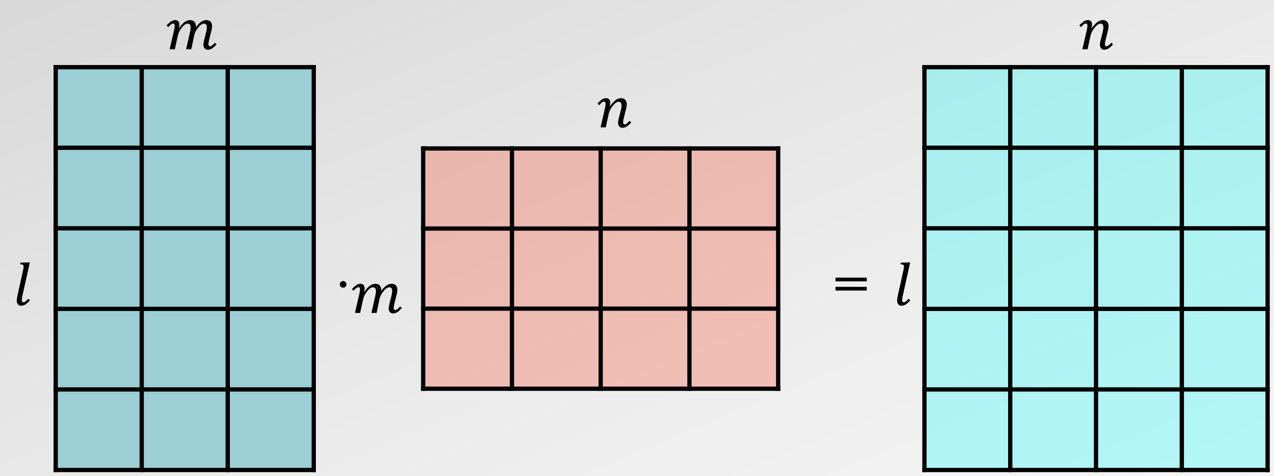
A more convenient way to write matrices and vectors

If $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$,

the *transpose* of **A** is found by interchanging the values across the diagonal, $a_{ij} \leftrightarrow a_{ji}$, which yields

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

Transpose of a matrix



$$\mathbf{AB} = [l \times m][m \times n] = [l \times n] = \mathbf{C}$$

The number of the columns in the first matrix **must** equal the number of rows in the second matrix. The result has the number of rows of the first and the number of columns of the second. If $\mathbf{AB} = \mathbf{C}$, then $C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}$ $i = 1$ to $l, j = 1$ to n.

This is the same expression we had for a 3x3 matrix with $l = 3$ and $n = 3$.

Non-square matrices

We will see many non-square matrices, e.g., 4 rows, 2 columns:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}$$

The transpose of **A** is obtained by reversing the order of the indices, i.e.,

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \end{bmatrix}$$

If **A** is a n x p matrix, then $\mathbf{A}^T$ is a p x n matrix

Non-square matrices

Consider the following

$$B = A^T A = \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^2 + a_{21}^2 + a_{31}^2 + a_{41}^2 & a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} + a_{41}a_{42} \\ a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} + a_{41}a_{42} & a_{12}^2 + a_{22}^2 + a_{32}^2 + a_{42}^2 \end{bmatrix}$$

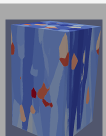
B is a symmetric 2x2 square matrix with: $\mathbf{B_{ij} = \sum_{k=1}^{4} A_{ki} A_{kj} \ i = 1..2 \ j = 1..2}$

$$\mathbf{AA}^T = \mathbf{C} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} & a_{11}a_{31} + a_{12}a_{32} & a_{11}a_{41} + a_{12}a_{42} \\ a_{11}a_{21} + a_{12}a_{22} & a_{21}^2 + a_{22}^2 & a_{21}a_{31} + a_{22}a_{32} & a_{21}a_{41} + a_{22}a_{42} \\ a_{11}a_{31} + a_{12}a_{32} & a_{21}a_{31} + a_{22}a_{32} & a_{31}^2 + a_{32}^2 & a_{31}a_{41} + a_{32}a_{42} \\ a_{11}a_{41} + a_{12}a_{42} & a_{21}a_{41} + a_{22}a_{42} & a_{31}a_{41} + a_{32}a_{42} & a_{41}^2 + a_{42}^2 \end{bmatrix}$$

C is a symmetric 4x4 square matrix with $C_{ij} = \sum_{k=1}^{2} A_{ki}A_{jk} \quad i = 1 \text{ to } 4 \quad j = 1 \text{ to } 4$

- Let $\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \end{bmatrix}$
- $\mathbf{A} + \mathbf{B} = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} & A_{13} + B_{13} \\ A_{21} + B_{21} & A_{22} + B_{22} & A_{23} + B_{23} \end{bmatrix}$
- $k\mathbf{A} = \begin{bmatrix} kA_{11} & kA_{12} & kA_{13} \\ kA_{21} & kA_{22} & kA_{23} \end{bmatrix}$
- $\mathbf{AB} = [l \times m][m \times n] = [l \times n] = \mathbf{C}$
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
- $(\mathbf{ABC})^T = ((\mathbf{AB})\mathbf{C})^T = \mathbf{C}^T(\mathbf{AB})^T = \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T$
- $(\mathbf{A}^T)^T = \mathbf{A}$



Some properties of matrices

- if $\mathbf{AB} = \mathbf{BA}$, they are said to *commute*. This only happens for very specific matrices. In general, $\mathbf{AB} \neq \mathbf{BA}$.

- $(\mathbf{BA})\mathbf{x} = \mathbf{B}(\mathbf{Ax}) = \mathbf{BAx}$ ($\mathbf{BA}$ is a composite map)

- $det\mathbf{AB} = det\mathbf{BA} = det\mathbf{A}det\mathbf{B}$ (scalars commute)

- only square matrices have inverses

- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

- $(\mathbf{ABC})^{-1} = ((\mathbf{AB})\mathbf{C})^{-1} = \mathbf{C}^{-1}(\mathbf{AB})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$

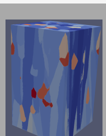
Some properties of matrices

We can write the dot product of two column vectors x and y as $\boldsymbol{x} \cdot \boldsymbol{y}$, in which $\boldsymbol{x}^T$ is the row vector obtained by transposing:

if $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$, then $\mathbf{x}^{\mathrm{T}} = [\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_4, \mathrm{x}_5]$.

Note: $$\mathbf{x} \cdot \mathbf{y} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i y_i} (= \mathbf{x_i y_i})$$

We will write expressions such as as $\mathbf{x} \cdot (\mathbf{Ay})$ as $\mathbf{x}^{\mathbf{T}}(\mathbf{Ay})$



The dot product

"In mathematics, **matrix calculus** is a specialized notation for doing multivariable calculus, especially over spaces of matrices. It collects the various partial derivatives of a single function with respect to many variables, and/or of a multivariate function with respect to a single variable, into vectors and matrices that can be treated as single entities. This greatly simplifies operations such as finding the maximum or minimum of a multivariate function and solving systems of differential equations. The notation used here is commonly used in statistics and engineering, while the tensor index notation is preferred in physics.

Two competing notational conventions split the field of **matrix calculus** into two separate groups. The two groups can be distinguished by whether they write the derivative of a scalar with respect to a vector as a column vector or a row vector. Both of these conventions are possible even when the common assumption is made that vectors should be treated as column vectors when combined with matrices (rather than row vectors). A single convention can be somewhat standard throughout a single field that commonly uses matrix calculus (e.g. econometrics, statistics, estimation theory and machine learning). However, even within a given field different authors can be found using competing conventions. Authors of both groups often write as though their specific convention were standard. Serious mistakes can result when combining results from different authors without carefully verifying that compatible notations have been used. Definitions of these two conventions and comparisons between them are collected in the layout conventions section."

An eigenvector $\mathbf{v}$ of $\mathbf{A}$ has the property
$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$

In other words, $\mathbf{v}$ does not change direction, even as the basis transforms under $\mathbf{A}$. The length of $\mathbf{v}$ after the transformation is λ times the original length (negative λ indicates that the direction has been reversed).

$\mathbf{v}$ is called an eigenvector and λ its eigenvalue.

It is more convenient to write the equation as $\mathbf{A}\mathbf{v} = \lambda\mathbf{I}\mathbf{v}$, where $\mathbf{I}$ is the identity matrix.

We thus can write: $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$.

As we discussed above, we can solve this equation using $det(\mathbf{A} - \lambda\mathbf{I}) = 0$

Suppose we have a known *N* x *N* matrix $\mathbf{A}$.

We solve $det(\mathbf{A} - \lambda\mathbf{I}) = 0$ for the *N* constants λ (the eigenvalues) and *N* vectors $\mathbf{v}$ (the eigenvectors).

Consider the matrix $\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 4 & 3 \end{bmatrix}$

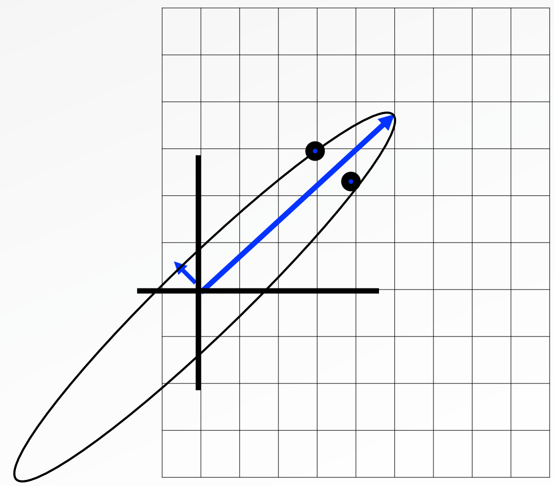
Solving for the eigenvalues, we have $det(\mathbf{A} - \lambda\mathbf{I}) = det\begin{bmatrix} 3-\lambda & 4 \\ 4 & 3-\lambda \end{bmatrix} = 0$

We find $\lambda = 7, -1$ with eigenvectors $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, respectively

The *eigenvalues* are the lengths of the semimajor and seminominor axes of an ellipse that includes the points.

The *eigenvectors* align the ellipse along the data.

Note that the eigenvectors make an orthogonal set, i.e., a rotated coordinate system along the directions of the greatest variance.



Interpretation of eigenvectors/values

We will have an equation that looks like $\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X} = \mathbf{Y}^T\mathbf{X}$ and will want $\hat{\boldsymbol{\beta}}$, knowing $\mathbf{X}$ and $\mathbf{Y}$. In the context of linear regression, $\mathbf{X}$ is a 2xN matrix made up from the x values and $\mathbf{Y}$ is a vector of y-values.

How will we solve it?

Taking the transpose of both sides of the equation, we have

$\left(\hat{\boldsymbol{\beta}}^T\mathbf{X}^T\mathbf{X}\right)^T = \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}$ and $(\mathbf{Y}^T\mathbf{X})^T = \mathbf{X}^T\mathbf{Y}$, or $\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$

Finally, we have: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$.

Knowing $\mathbf{X}$ and $\mathbf{Y}$, we can easily find $\hat{\boldsymbol{\beta}}$.

Solving for the coefficients 1: least squares